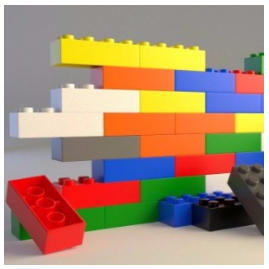
Oct 13, 2017 Version 3

Assembly with Megahit V.3



DOI
dx.doi.org/10.17504/protocols.io.kancsde

James E Thornton Jr[1]

[1]Hurwitz Lab

Metafunc Course 2017

James E Thornton Jr
Hurwitz Lab

OPEN ACCESS

DOI: dx.doi.org/10.17504/protocols.io.kancsde

Protocol Citation: James E Thornton Jr 2017. Assembly with Megahit. **protocols.io** **https://dx.doi.org/10.17504/protocols.io.kancsde**

Protocol status: Working

Created: October 13, 2017

Last Modified: March 28, 2018

Protocol Integer ID: 8238

Abstract

Co-assembly using Megahit.

1 Log into the HPC.

Command

```
$ ssh hpc
$ ocelote
```

2 From your home directory, open .bashrc file for editing.

Command

```
$ nano .bashrc
```

Note

Remember, you are already in your home directory after logging into ocelote.

3 Input the following line into your .bashrc file:

Command

This will allow us to execute tools found in /rsgrps/bh_class/bin without specifying the path name.

```
export PATH=/rsgrps/bh_class/bin:$PATH
```

4 Save and close the .bashrc file

5 Move into your project directory.

Command

```
$ cd /rsgrps/bh_class/username
```

6 Create a directory for assembly output. Then move into that directory.

Command

```
$ mkdir assembly
$ cd !$
```

7 Make a directory for fasta files. This is the format of your files after doing quality control. You will also need to move all fasta files into this directory.

Command

```
mkdir fasta
mv fastq/*.fasta fasta
```

Note

This step assumes that the fasta files are in the fastq directory. Alter the command accordingly if this is not the case.

8 Make directories for standard out and standard error.

Command

```
mkdir std-out std-err
```

9 Before we continue, determine if you have single end or paired end files. If you have two files per SRR number, you have paired end reads. Otherwise, you have single end reads.

1. If you have single end reads proceed to step 10.
2. If you have paired end reads, skip to step 11.

10 Assembly script for SINGLE END FILES

Create a script called run-assembly.sh

Command

```
#!/bin/bash
#PBS -W group_list=bh_class
#PBS -q windfall
#PBS -l select=1:ncpus=20:mem=40gb
#PBS -l pvmem=38gb
#PBS -l walltime=24:00:00
#PBS -l cput=48:00:00
#PBS -M netid@email.arizona.edu
#PBS -m bea
FASTA_DIR='/rsgrps/bh_class/username/fasta'
ASSEM_DIR='/rsgrps/bh_class/username/assembly'
MIN_CONTIG_LEN=500
OUT_DIR='/rsgrps/bh_class/username/assembly/megahit-out'
cd $ASSEM_DIR
SINGLEs=`ls $FASTA_DIR/*.fasta | python -c 'import sys; print
```

Note

OUT_DIR does NOT need to be created prior to running this script. Megahit will make the directory on its own.

11 Assembly script for PAIRED END FILES

Create a script called run-assembly.sh

Command

```
#!/bin/bash
#PBS -W group_list=bh_class
#PBS -q windfall
#PBS -l select=1:ncpus=20:mem=40gb
#PBS -l pvmem=38gb
#PBS -l walltime=24:00:00
#PBS -l cput=48:00:00
#PBS -M netid@email.arizona.edu
#PBS -m bea
FASTA_DIR='/rsgrps/bh_class/username/fasta'
ASSEM_DIR='/rsgrps/bh_class/username/assembly'
MIN_CONTIG_LEN=500
OUT_DIR='/rsgrps/bh_class/username/assembly/megahit-out'
cd $ASSEM_DIR
R1s=`ls $FASTA_DIR/*_1.fasta | python -c 'import sys; print
```

Note

OUT_DIR does NOT need to be created prior to running this script. Megahit will make the directory on its own.

12 Run the assembly:

Command

```
$ chmod +x run-assembly.sh
$ qsub -e std-err/ -o std-out/ run-assembly.sh
```

13 You can check the status of your job with the following command:

Command

```
$ qstat -u username
```

Note

Job runtime will vary depending on the size of your dataset.

14 Upon job completion, get assembly statistics using MetaQuast on CyVerse.

Protocol

NAME

Assembly Stats with MetaQuast

CREATED BY

James E Thornton Jr

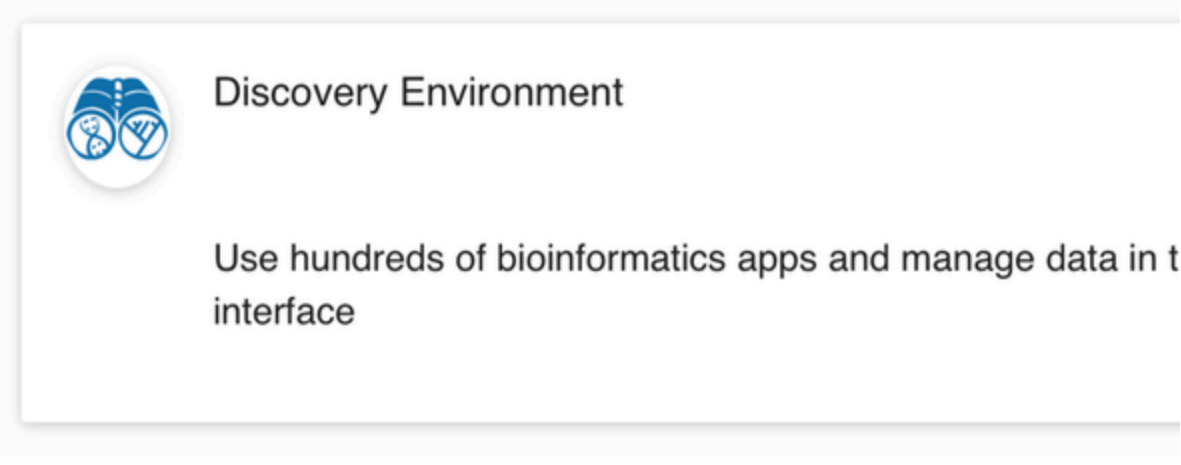
PREVIEW

14.1 Go to https://user.cyverse.org/
https://user.cyverse.org/

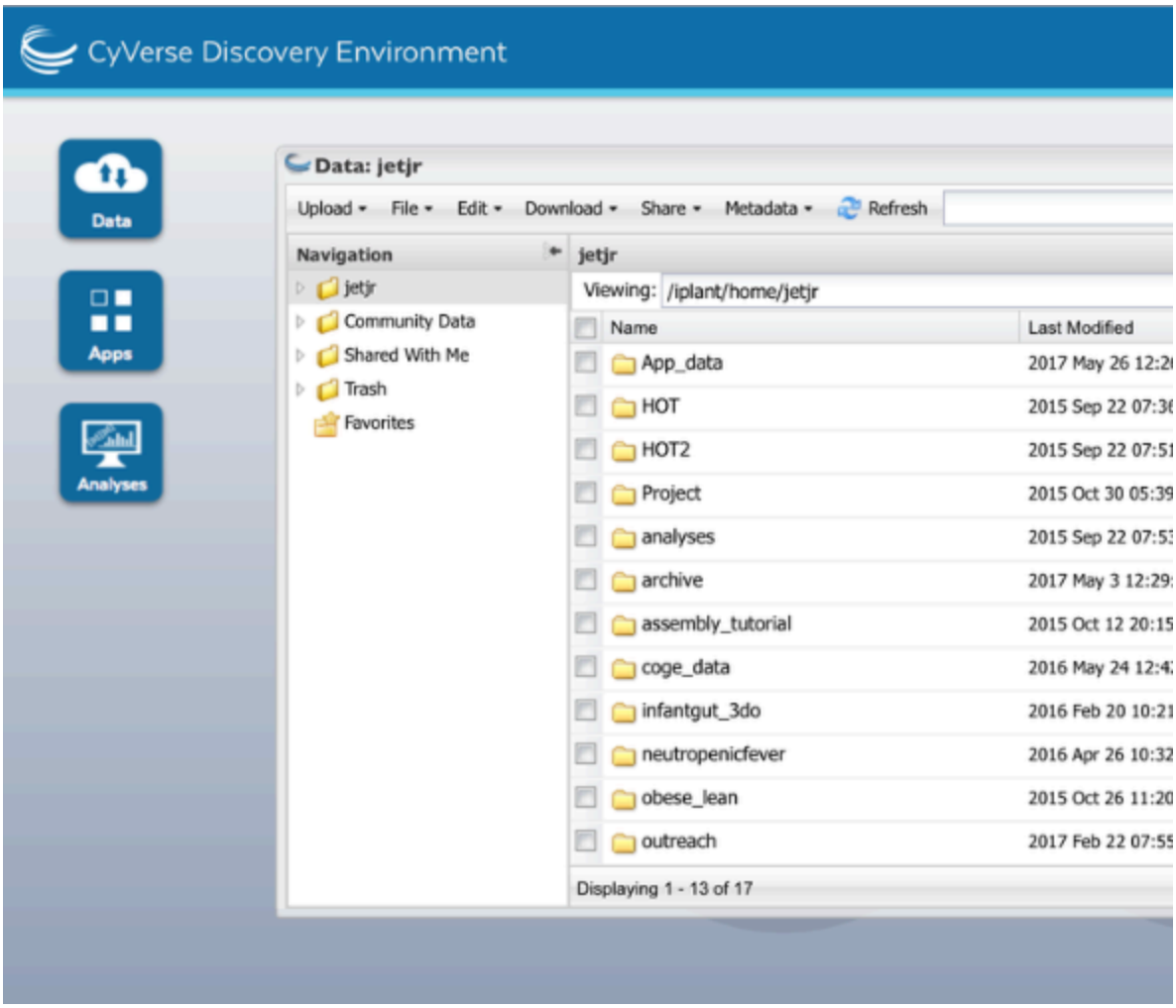
14.2 Click "Sign Up" to create an account.

14.3 After account creation go back to https://user.cyverse.org/ and login with your account.
https://user.cyverse.org/

14.4 Launch the discovery environment.



14.5 Click the "Data" button found on the left. Navigate to your user folder.



5

14.6 Click "Upload" > "Simple Upload From Desktop"

14.7 Upload your final.contigs.fa file generated from Megahit.

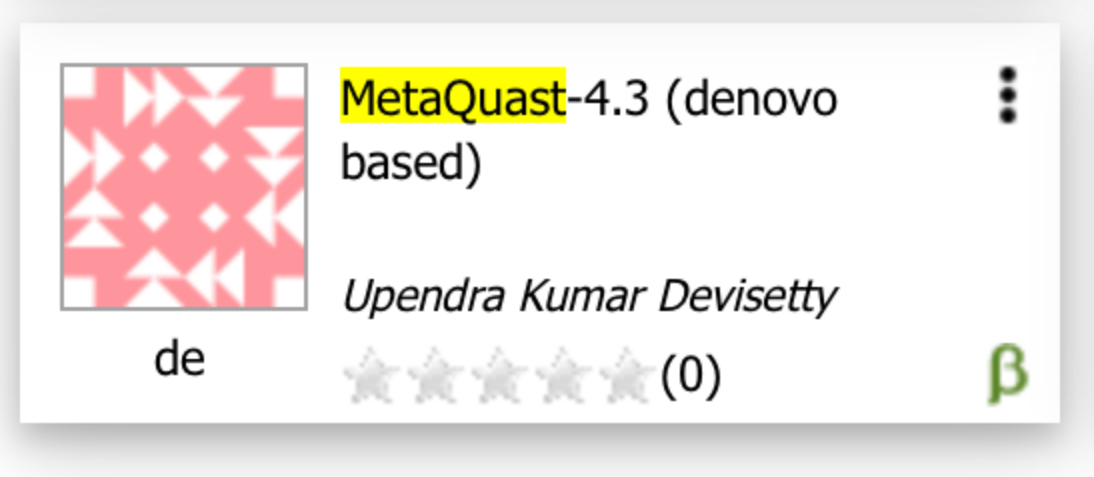
Note

Important: You must scp your contigs to your local machine before you can upload.

$ scp username@sftp.hpc.arizona.edu:/rsgrps/bh_class/username/assembly/megahit-out/final.contigs.fa .

14.8 Once your upload is complete, click on the "Apps" button found on the left.

14.9 Search for "MetaQuast". Click on MetaQuast-4.3 (denovo based)



14.10 Under the "Fasta file(s)" tab, select the newly uploaded final.contigs.fa file. This is the only parameter that needs to change. Click "Launch Analysis".

14.11 Once MetaQuast is complete (email notification), navigate to the output found in the "analyses" folder in your data storage.

14.12 Download the "report.html" file found in the MegaQuast output folder.

14.13 Open the report.html file to see a summary of assembly statistics.